December 10, 2018

Visualizing classifier performance

Lori Levy

John Burge

Visualizing classifier performance

ABSTRACT

Classifiers are characterized by many measures, e.g., precision, recall, accuracy, sensitivity, specificity, receiver operating curve (ROC), fall-out, class imbalance, etc. Some measures are synonymous, e.g., sensitivity and recall. The profusion of terms causes some difficulties.

This disclosure leverages the observation that most classifier performance metrics derive from just four elemental quantities, e.g., true positives, false positives, true negatives, and false negatives. Disclosed is a simple graphic that displays the four elemental quantities and enables quick, visual, estimation of the remaining measures of classifier performance.

KEYWORDS

- classifier performance
- information visualization
- visual presentation
- receiver operating curve
- class imbalance

BACKGROUND

Classifiers are characterized by many measures, e.g., precision, recall, accuracy, sensitivity, specificity, receiver operating curve (ROC), fall-out, class imbalance, etc. Some measures are synonymous, e.g., sensitivity and recall. The profusion of terms causes some difficulties. Yet all the terms are derived from just four elemental quantities, e.g., the numbers of true positives, false positives, true negatives, and false negatives.

For example,

- sensitivity, also known as recall, hit-rate, true positive rate, etc., is defined as

$$\frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

- specificity, also known as selectivity, true negative rate, etc., is defined as

$$\frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}$$

- precision, also known as positive predictive value, is defined as

$$\frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

- miss-rate, also known as false negative rate, is defined as

$$\frac{\text{false negatives}}{\text{true positives} + \text{false negatives}}$$

- fall-out, also known as false positive rate or false alarm rate, is defined as

$$\frac{\text{false positives}}{\text{true negatives} + \text{false positives}}$$

- false discovery rate is defined as

$$\frac{\text{false positives}}{\text{true positives} + \text{false positives}}$$

- accuracy is defined as

$$\frac{\text{true positives} + \text{true negatives}}{\text{true positives} + \text{true negatives} + \text{false positives} + \text{false negatives}}$$

- receiver operating curve is defined as a plot of recall (true positive rate) against fall-out (false positive rate).

DESCRIPTION

Per the techniques of this disclosure, ground truth and classifier results are each represented as binary-colored concentric circles such that overlapping arcs represent the four elemental quantities. The resulting visual representation enables easy evaluation of various performance metrics.

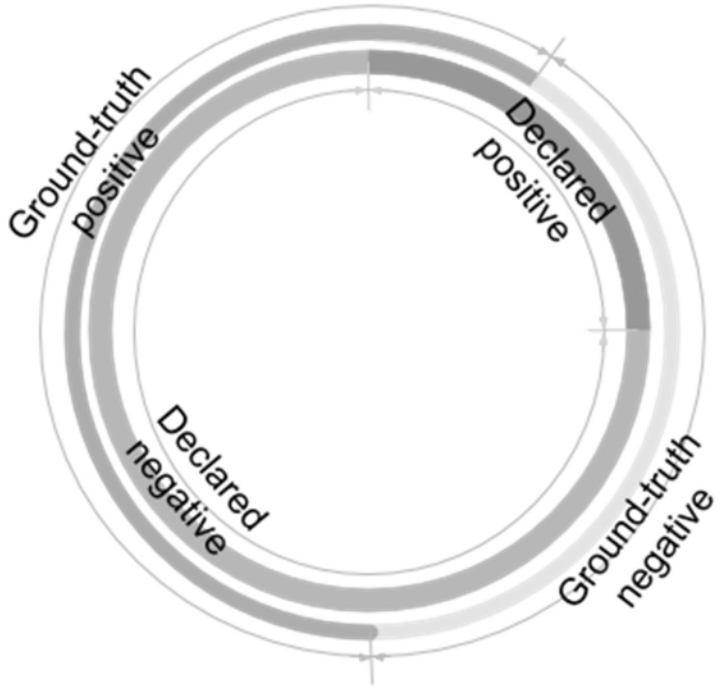

Fig. 1: Ground truth and classifier results represented as overlapping concentric circles

This is illustrated in Fig. 1, where the green arc represents ground-truth positive data, the yellow arc represents ground-truth negative data, the blue arc represents positive data as declared by the classifier, and the grey arc represents negative data as declared by the classifier. The size of each arc is based on a quantity for the classification.

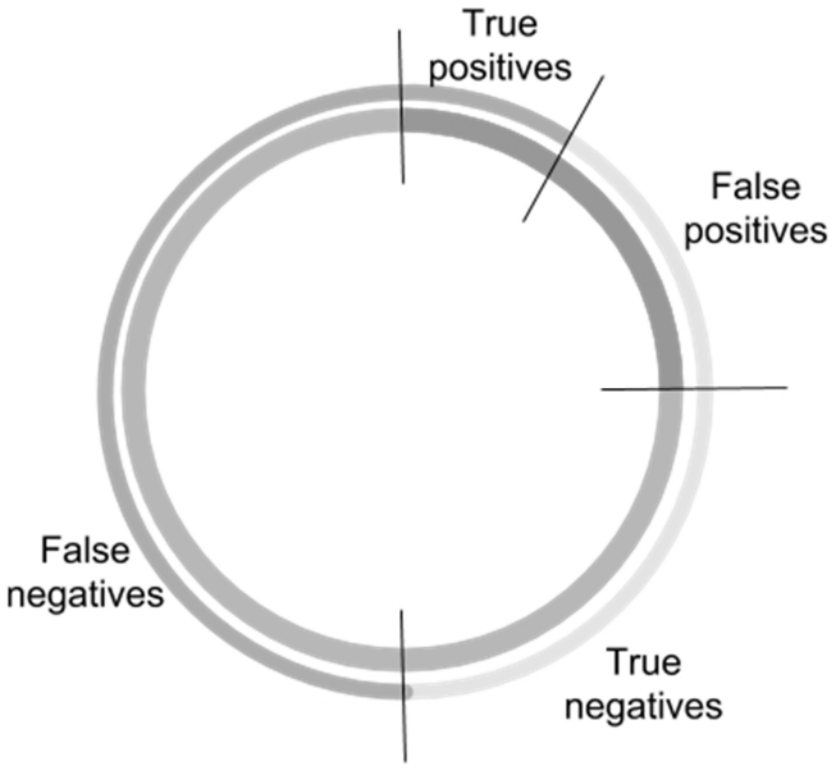

Fig. 2: The four elemental quantities represented as overlaps of arcs

With the convention established by Fig. 1, the following observations, shown in Fig. 2, hold:

- True positives are represented by the overlap between the blue and the green arcs.
- False positives are represented by the overlap between the blue and the yellow arcs.
- False negatives are represented by the overlap between the grey and green arcs.
- True negatives are represented by the overlap between the grey and the yellow arcs.

Further, it holds that:

- Sensitivity equals the ratio of the blue-green overlap to green arc.
- Specificity equals the ratio of the yellow-grey overlap to the yellow arc.
- Precision equals the ratio of the blue-green overlap to the blue arc.
- Miss-rate equals the ratio of grey-green overlap to the green arc.

- Fall-out equals the ratio of blue-yellow overlap to the yellow arc.
- False discovery rate equals the ratio of grey-green overlap to the green arc.
- Class imbalance is the ratio of the green arc to the yellow arc.

Other commonly used performance metrics can similarly be worked out as ratios of overlaps to arcs. In this manner, a relatively simple diagram such as that of Fig. 2 conveys almost all of the numerous performance metrics typically used to characterize a classifier.

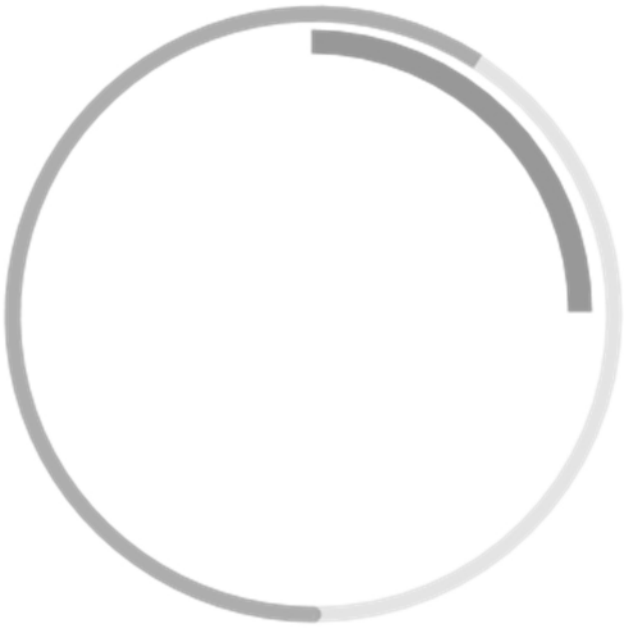

Fig. 3: Simplified visual representation of classifier performance

Fig. 3 shows a further simplified, hence minimalist, representation of classifier performance. This representation is derived by omitting the grey arc, with the understanding that the grey arc invisibly occupies the complement of the blue arc. The representation of Fig. 3 is further free of clutter, and still conveys the same information as that conveyed by Fig. 2.

(a) (b)

Fig. 4: Different visual representations of classifier performance

Variations of the visual representations disclosed herein are possible, e.g., interchanging the positions of ground-truth and classifier-result circles (Fig. 4a), representing ground-truth and classifier-result data linearly (Fig. 4b), in other geometric shapes such as square or other regular polygons, etc.

In this manner, the techniques of this disclosure convey effectively several performance metrics of a classifier with a minimal cognitive load, e.g., just three-to-four colors and properly oriented overlap of concentric circles. The visualizations can be incorporated in any product that displays classifications.

CONCLUSION

This disclosure leverages the observation that most classifier performance metrics, e.g., precision, recall, accuracy, sensitivity, specificity, receiver operating curve (ROC), fall-out, class imbalance, etc., derive from four elemental quantities, e.g., true positives, false positives, true negatives, and false negatives. Using a simple graphic that displays the four elemental quantities, the techniques of this disclosure enable a quick, visual estimation of the measures of classifier performance with minimal cognitive load.